NMR structure of cysteinyl-phosphorylated enzyme IIB of the N,N '-diacetylchitobiose-specific phosphoenolpyruvate-dependent phosphotransferase system of Escherichia coli

AB, E; Schuurman-Wolters, G.K.; Nijlant, D.; Dijkstra, K.; Saier, M.H.; Robillard, G.T.; Scheek, R.M.

Published in:
Journal of Molecular Biology

DOI:
10.1006/jmbi.2001.4623

Publication date:
2001

doi:10.1006/jmbi.2001.4623 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **308**, 993–1009

JMB

NMR Structure of Cysteinyl-phosphorylated Enzyme IIB of the *N,N′*-diacetylchitobiose-specific Phosphoenolpyruvate-dependent Phosphotransferase System of *Escherichia coli*

Eiso AB[1], Gea K. Schuurman-Wolters[1], Dieter Nijlant[1], Klaas Dijkstra[1] Milton H. Saier[2], George T. Robillard[1] and Ruud M. Scheek[2]*

[1]*The Groningen Biomolecular Science and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 4 9747 AG Groningen, The Netherlands*

[2]*Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA*

**Corresponding author*

The determination by NMR of the solution structure of the phosphorylated enzyme IIB (P-IIB^{Chb}) of the *N,N′*-diacetylchitobiose-specific phosphoenolpyruvate-dependent phosphotransferase system of *Escherichia coli* is presented. Most of the backbone and side-chain resonances were assigned using a variety of mostly heteronuclear NMR experiments. The remaining resonances were assigned with the help of the structure calculations.

NOE-derived distance restraints were used in distance geometry calculations followed by molecular dynamics and simulated annealing protocols. In addition, combinations of ambiguous restraints were used to resolve ambiguities in the NOE assignments. By combining sets of ambiguous and unambiguous restraints into new ambiguous restraints, an error function was constructed that was less sensitive to information loss caused by assignment uncertainties. The final set of structures had a pairwise rmsd of 0.59 Å and 1.16 Å for the heavy atoms of the backbone and side-chains, respectively.

Comparing the P-IIB^{Chb} solution structure with the previously determined NMR and X-ray structures of the wild-type and the Cys10Ser mutant shows that significant differences between the structures are limited to the active-site region. The phosphoryl group at the active-site cysteine residue is surrounded by a loop formed by residues 10 through 16. NOE and chemical shift data suggest that the phosphoryl group makes hydrogen bonds with the backbone amide protons of residues 12 and 15. The binding mode of the phosphoryl group is very similar to that of the protein tyrosine phosphatases. The differences observed are in accordance with the presumption that IIB^{Chb} has to be more resistant to hydrolysis than the protein tyrosine phosphatases. We propose a proton relay network by which a transfer occurs between the cysteine SH proton and the solvent *via* the hydroxyl group of Thr16.

Keywords: PTS·IIB^{Chb}; IIB-chitobiose; phosphocysteine; cysteinyl-phosphate

Abbreviations used: C8E5, *n*-octyl-penta(ethylene glycol); cel, cellobiose; chb, *N,N′*-diacetylchitobiose; DDD, distance bounds driven dynamics; EI, EII, enzyme-I, enzyme-II; HPr, histidine-containing protein; IIA^{Chb}, IIB^{Chb}, IIC^{Chb}, enzyme IIA, IIB and IIC of the *N,N′*-diacetylchitobiose PTS (respectively); MD, molecular dynamics; NOE, nuclear Overhauser effect; PEP, phosphoenolpyruvate; PTPase, protein tyrosine phosphatase; PTS, phosphoenolpyruvate-dependent phosphotransferase system; SA, simulated annealing; TAG, Tris-acetate/glycerol; TPPI, time-proportional phase incrementation; DSS, 2,2-dimethyl-2-silapentase-5-sulfonate.

E-mail address of the corresponding author: scheek@chem.rug.nl

0022-2836/01/050993–17 $35.00/0

Introduction

Enzyme IIB^{Chb} of *Escherichia coli* is the central energy-coupling domain of the phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS) of *E. coli* for the carbohydrate *N*,*N*′-diacetylchitobiose ((GlcNac)$_2$). The *chb* operon of *E. coli* was originally named *cel* operon and was thought to be a cryptic operon coding for the enzymes of the cellobiose PTS.[1,2] Keyhani & Roseman[3] found that the operon is normally inducible and codes for the *N*,*N*′-diacetylchitobiose PTS. Here, we use the new *chb* nomenclature.

After transport from the periplasmic to the cytoplasmic face of the membrane by enzyme IIC^{Chb}, the carbohydrate is phosphorylated by IIB^{Chb} at the expense of PEP. Enzyme IIB^{Chb} receives its phosphoryl group *via* a series of phosphorylation and dephosphorylation reactions involving enzyme I, HPr and finally IIA^{Chb}. For general reviews on the PEP-dependent PTS we refer to Postma *et al.*,[4] Meadow *et al.*,[5] Lolkema *et al.*,[6] Saier & Reizer,[7] Robillard & Broos.[8] In the case of the mannitol PTS, a conformational coupling was proposed between the phosphorylation-state of the IIB domain of the IICBA protein and transport of the carbohydrate through the IIC domain.[9–11] *E. coli* IIB^{Chb} is a protein of 106 amino acid residues and differs in this respect from IIB domains for most other carbohydrates, which usually are integral parts of multi-domain proteins. IIB^{Chb} has been shown to be phosphorylated at residue Cys10.[12] The crystal structure of the wild-type and the solution structure of the Cys10Ser mutant were solved previously.[13,14] The work presented here describes the structure of the wild-type enzyme, which carries a phosphoryl group at the active-site residue Cys10.

The only other classes of enzymes where phosphorylation is known to occur on a cysteine residue are the protein tyrosine phosphatases (PTPases)[15,16] and dual specificity (Tyr/Thr) protein phosphatases.[17,18] In both cases, as well as in IIB^{Chb}, the active cysteine lies in a strand-loop-helix motif at the end of the β-strand, and the phosphoryl group is stabilized by backbone amide hydrogen bond donors.

Results

Phosphorylation of IIB^{Chb}

Under the conditions of high buffer capacity used (see Materials and Methods), the samples of phosphorylated IIB^{Chb} remained completely phosphorylated for approximately one week. This was inferred by monitoring the presence of resonances specific for the unphosphorylated form of the enzyme in a ^{15}N-HSQC. The samples could be regenerated by dialysis against fresh buffer, but the spectra of the resulting samples showed a considerable number of additional resonances compared with a fresh sample. Peak shifts from the pH instability, caused by the release of pyruvate, were generally limited, but most pronounced for the backbone resonances of His5.

Resonance assignments

Backbone and side-chain assignments

Most of the assignments for the backbone resonances were readily obtained, except for the Ser15 H^N resonance, which could be detected only in NOE traces of ^{15}N and ^{13}C edited 3D-NOESY and 2D-NOESY spectra. Signals at 10.7 ppm from Ser15 H$^\alpha$,H$^{\beta1}$,H$^{\beta2}$ and Met14 H$^\alpha$ were found, to which Ser15 H^N could be assigned. Lys3 H^N appears to be degenerate. We observed two sets of resonances that could be assigned to Lys3 and are both consistent with the Lys4 assignments in all spectra. They differ significantly only in the chemical shift value of the amide proton. We presume that this is caused by partial hydrolysis of the N-terminal methionine residue. Because in this study we focussed on the active-site structure specifically, and Lys3 is located at the opposite side of the enzyme, we did not pursue the assignment and conformation of this residue in further detail. To avoid distortion of the structure by inconsistent restraints, NOEs that were specifically assigned to the resonances of the Lys3 with the lower field amide chemical shift were not used for generating distance restraints.

The backbone experiments aimed at observation of the β-resonances, viz. the HBHA(CBCA)(CO)NH, CBCA(CO)NH, HCACBCO together with the 3D-HCCH-TOCSY, the C$^\beta$H$^\delta$ and C$^\beta$H$^\varepsilon$ for the aromatic resonances, and the ^{15}N-TOCSY-HSQC, were not sufficient to complete the side-chain resonance assignments. Most of the remaining side-chain resonances were assigned during the structure calculations from the ^{13}C-HSQC-NOESY, by searching for unexplained cross-peak intensity at the resonance frequencies of nearby nuclei in the following manner. Planes perpendicular to the ω3 domain were taken corresponding to the frequencies of assigned protons expected to show intra-residue and sequential NOEs. Often inspection of a contour plot of the resulting data, sometimes combined with NOE-intensity and chemical shift information, allowed many of the remaining side-chain resonances to be assigned directly. The methyl resonances of methionine residues 14, 23 and 63 were assigned from the NOE spectra alone, based on their typical ^{1}H and ^{13}C chemical shift values and contacts to nearby nuclei.

Currently we have assignments for all nuclei except Val32 C$^{\gamma2}$ (possible overlap with C$^{\gamma1}$), Ser21 H$^{\beta2}$ (possible overlap with H$^{\beta1}$), Ser81-C$^{\beta1}$ (possible overlap with C$^\alpha$) and Met1 H$^\gamma$,C$^\gamma$, and exchangeable protons. The chemical shifts are deposited at the BioMagResBank (see Materials and Methods).

Shift differences between Cys10Ser-IIB^{Chb} and P-IIB^{Chb}

The differences between the Cys10Ser mutant and phospho-IIB^{Chb} backbone chemical shifts are shown in Figure 1(a). The largest differences occur in the neighborhood of the active-site region: residues 9-16, 39-44, 59-60 and 84. For the remaining residues, the chemical shifts, and consequently the structure in solution, are very similar. The resonances of His5 also show a larger shift, but this is more likely caused by the high sensitivity of this residue to pH instabilities during the measurements.

The exact changes in chemical shifts upon phosphorylation of the wild-type enzyme are not known, since the resonance assignments of the wild-type enzyme are not available. However, from the differences between the ^{15}N-HSQC spectra shown in Figure 1(b), we can conclude that they will be of the same order of magnitude as the differences between the phosphorylated wild-type and the Cys10Ser mutant.

Structure determination

NOE assignments

The number of NOE intensities collected and the resulting number of restraints generated for each structure calculation or refinement protocol are listed in Table 1. We collected as many peaks as possible from the ^{13}C and ^{15}N edited NOESY spectra, with the exception of peaks from some very crowded spectral regions. These usually do not contain much extra information because of the high level of ambiguity and the uncertainty in the peak positions and/or resonance assignments. The number of restraints used for each residue is displayed in Figure 2.

Structure calculations

The use of wide (square well potential) distance bounds for restraining can easily obscure information present in small differences between NOEs. These can nevertheless be significant, for example if they occur in the same relaxation frame. Therefore we prefer to use tight bounds with a relatively low force constant. We noticed that local convergence to a low target function is somewhat better in the searching phase if tight restraints are used. It also makes comparison of experimental and calculated data more straightforward. However, the resulting set of structures will likely be biased towards low rmsd values.

In the final round of structure calculations, all unambiguous restraints were used in an embedding procedure to generate 32 sets of 4D coordinates. These were subsequently further refined in 4D and 3D against all available restraint data using distance bounds driven dynamics (DDD). Finally, several rounds of refinement were performed in explicit solvent using GROMACS.

Because we focussed on the region around the active site, from the 32 structures calculated in the last round eight with relatively low per-residue R_2^x values[19] for the residues in the active site (residues 10-16, 59, 84) were selected for further analysis. The average rmsd values of non-hydrogen atoms were 0.59 and 1.16 Å for the backbone atoms and side-chain atoms, respectively, and 0.38 and 0.88 Å for residues 3-100, respectively. Figure 3 shows the differences between the selected structures, as well as the differences between the representative structures and the wild-type and Cys10Ser structures.

Analysis of violations and quality

Table 2 lists energies, violation data and *R*-factors for the eight selected structures, the representative structure before and after free stee-

Table 1. Summary of NOEs and restraints used for the structure determination

NOE peaks		Total	Unassigned
^{13}C-HSQCNOESY		3303	511
^{15}N-NOESYHSQC		1207	357
2D-NOESY		327	-
Restraints		Total	Ambiguous
H bond		103	-
DG		2525	-
DDD		3605	1376
GROMACS	Set A	2197	853
	Set C+U	1886	1373
	Total	4083	2226

Assigned NOE peaks belong to one specific spin-pair, not necessarily stereospecifically assigned. For the unassigned peaks, all possible spin pairs within a cut-off distance were considered. Restraints containing ambiguities arising soley from rapidly rotating methyl groups and aromatic rings are considered unambiguous. In the DDD stage, the prochiralities were handled by flipping chiralities. In the GROMACS stage, the prochiralities were expanded into ambiguities (see Materials and Methods). The set of restraints denoted A refers to the restraints that originate from assigned NOE peaks. For this set, identical upper and lower bounds were used. The C + U set refers to the upperbound restraints originating from unassigned NOEs and from combinations of ambiguous restraints due to shared assignments.

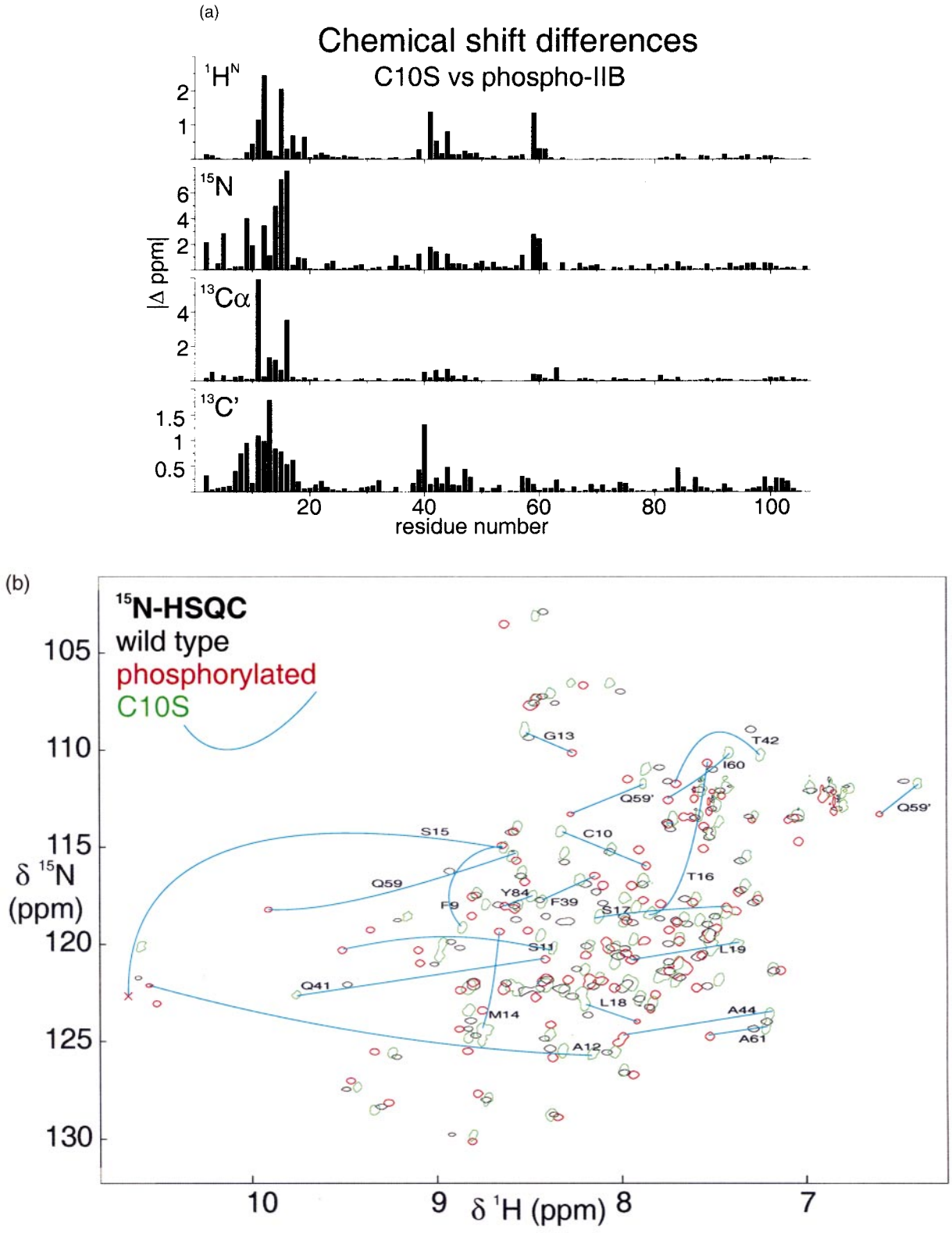

Figure 1. NMR data for enzymes IIB^{Chb}. (a) Backbone chemical shift differences between P-IIB^{Chb} and Cys10Ser-IIB^{Chb}. (b) ^{15}N-^{1}H-HSQC spectra of Cys10Ser-IIB^{Chb}, wild-type IIB^{Chb} and P-IIB^{Chb}. Large chemical shift differences between Cys10SerIIB^{Chb} and P-IIB^{Chb} in the vicinity of the active site are indicated with light-blue curves. The cross at 10.7 ppm indicates the position of the unobserved S15 resonance in P-IIB^{Chb}.

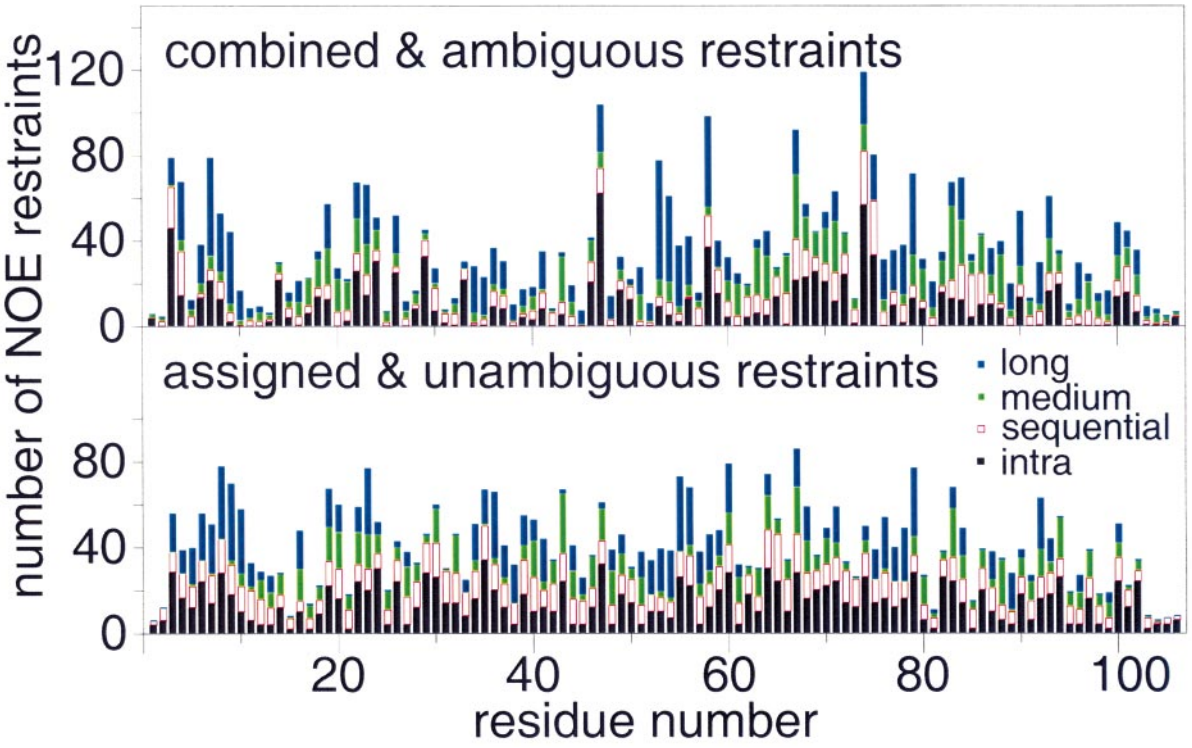

Figure 2. Number of restraints per residue, used in the GROMACS simulated annealing runs. In the upper panels the restraints originating from unassigned NOE peaks and combinations of restraints are shown, these were used as upper bound restraints only. In the lower panel, the restraints originating from assigned NOE peaks are displayed. For these restraints, both upper and lower bounds were applied.

pest descents minimization are listed. The energies for all structures are comparable, indicating that the NOE restraints did not impose unreasonable strain on the structure. Also the R_2^x values and total violation vary only moderately.

The Ramachandran plot of the representative structure, shown in Figure 4, indicated a structure of good quality with 87.5 % of the phi/psi values in the most favoured regions, and none in the disallowed regions. The equivalent resolution values given by the program PROCHECK[20] were 1.4 Å based on the hydrogen bond energy standard deviation and maximally 1.1 Å based on the percentage residues in most favoured regions, the $\chi 1$ pooled standard deviation and standard deviation of $\chi 2$ *trans* angles.

Table 2. GROMACS energies of the final structures

	NMR-models $\langle E \rangle$ (σ) kJ/mol	WTP-R	WTP-F
Bonds	Constr.	Constr.	Constr.
Angles	654(35)	699	460
Proper dihedrals	479(22)	443	424
Improper dihedrals	105(17)	123	54
Lennard-Jones-14	708(12)	720	694
Lennard-Jones	−4426(44)	4374	−4501
Coulomb[a]	−6026(428)	−6372	−5695
Distance restraints			
Set A: upper bounds	953(37)	901	1114
Set A: lower bounds	894(18)	817	932
Set C + U: upper bounds	208(10)	210	294
⟨Energy/restraint⟩	0.50(0.01)	0.47	0.57
Sum of violations (Å)	713(23)	607	653
⟨Violation/restraint⟩ (Å)	0.176(0.005)	0.150	0.161
R_2^x	9.59(0.08)	9.23	9.74

Energies, *R*-factors and violations for the eight selected NMR structures, the representative model (WTP-R) and for the representative model after 55 steps of unrestrained steepest descents minimization (WTP-F). R_2^x values[19] were calculated with the set of assigned NOEs.

[a] Coulomb interactions are without solvent-solvent and solvent-protein contribution.

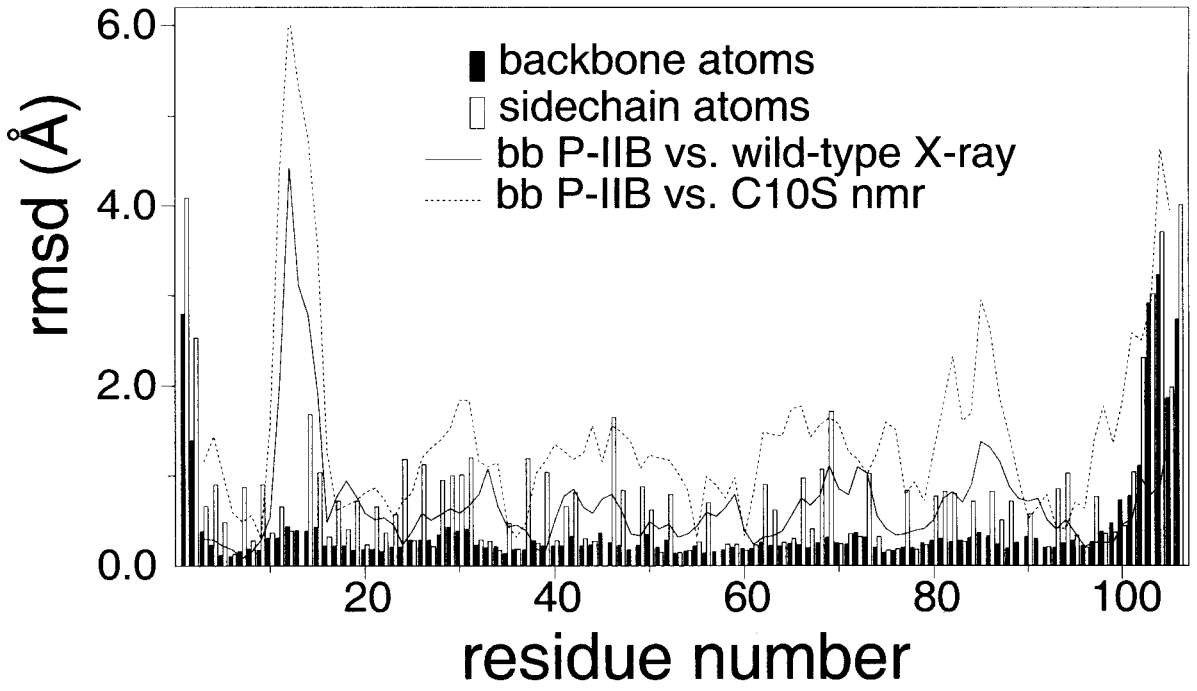

Figure 3. Differences between IIB^{Chb} structures. Positional rms differences of backbone atoms and heavy side-chain atoms for the eight selected structures of P-IIB^{Chb} (bars), and rms differences between the representative P-IIB^{Chb} structure and the wild-type X-ray structure and the mean structure of Cys10Ser-IIB^{Chb}, respectively (lines). The coordinates were superimposed on the C^{α} atoms of residues 3-100 for the superposition of the P-IIB^{Chb} structures, and on the C^{α} atoms of residues 3-10, 17-100 for the superposition of the representative P-IIB, wild-type X-ray and C10S structures.

To analyze the local correspondence with the data of the structures we used the assigned NOE-intensities to calculate sixth-root *R*-factors for each residue separately. The results are given in Figure 5. Except for the termini, for which the low number of restraints caused convergence problems, there are no extreme outliers.

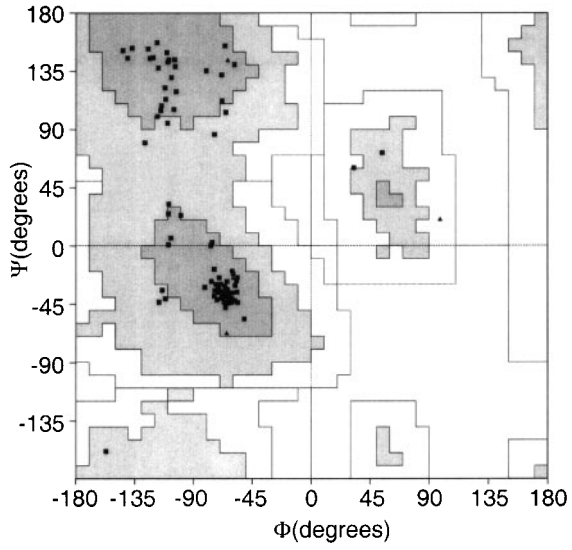

Figure 4. Ramachandran plot of the representative model of P-IIB^{Chb}, made using the program PROCHECK.[62] Glycine residues are represented by tri-

Figure 6 shows a comparison of the observed and calculated sixth-root NOE intensities for the eight selected structures. Outliers are easily noticed in such a plot. Approximately 99.9 % of the restraints correspond within 1.0 Å to the distances in the structures. The maximum disagreement per structure ranges from 1.5-2.0 Å. Relatively large discrepancies occur for the side-chain amide protons of Gln68 located at the surface of the protein, to side-chain protons of Leu64 and amide and side-chain protons of Lys74. In the models, the conformation of the Gln68 side-chain varies from the vicinity of 64 to that of 74. Calculating ensemble-averaged NOE intensities for two structures comprising both variants yielded significantly better correspondence with the observed intensities for this side-chain. Therefore, it is likely that the violations are caused by mobility of the Gln68 side-chain between the vicinity of residues 64 and 74, which cannot be modelled correctly in one structure.

In order to have an independent confirmation of the observed differences in the P-loop compared with the wild-type X-ray structure, ^{15}N-^{1}H dipolar coupling measurements were performed[21] We were able to measure dipolar couplings for the backbone amide groups of 73 residues in the oriented phase. We could measure isotropic

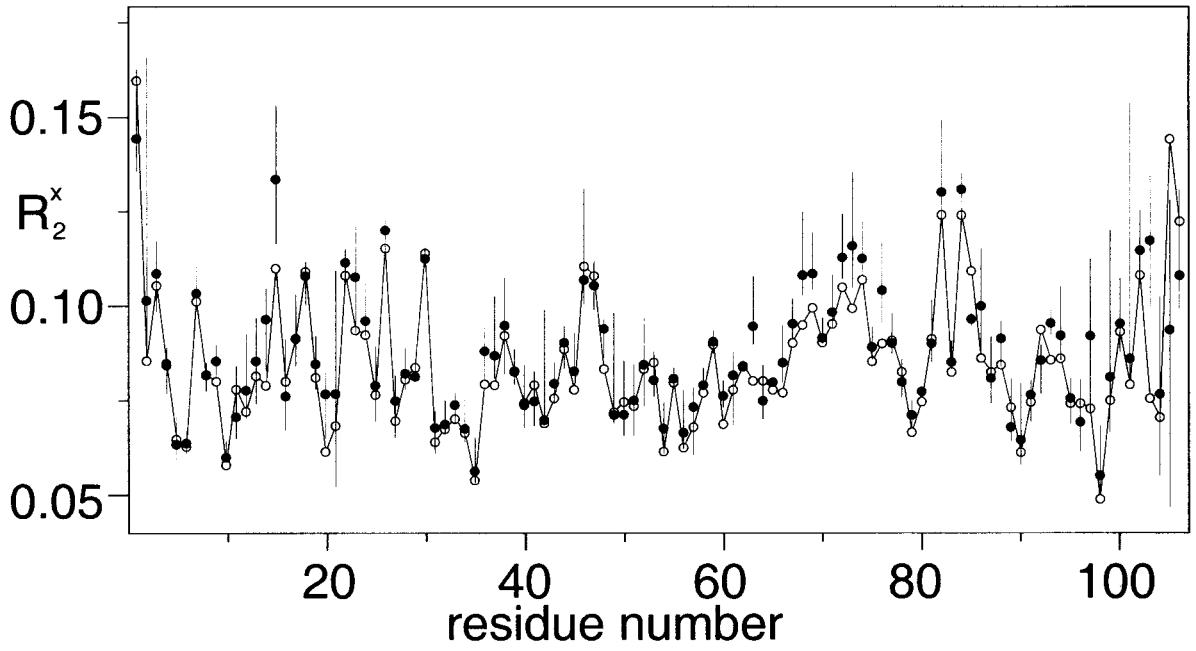

Figure 5. Per-residue R_2^x values of P-IIB^{Chb} NMR structures (total range, vertical lines; median, filled circle) and representative structure (connected open circles).

reference values for only 48 of those due to difficulty in keeping the sample phosphorylated at 298 K in the presence of the C8E5/octanol mixture. As the isotropic *J*-values have a rather narrow range, this has not hindered the interpretation of the data. An average isotropic value of 93.3 Hz was taken to calculate the contribution of the dipolar coupling if no measured reference value was available. The optimal molecular alignment tensor was determined for each model by minimizing the

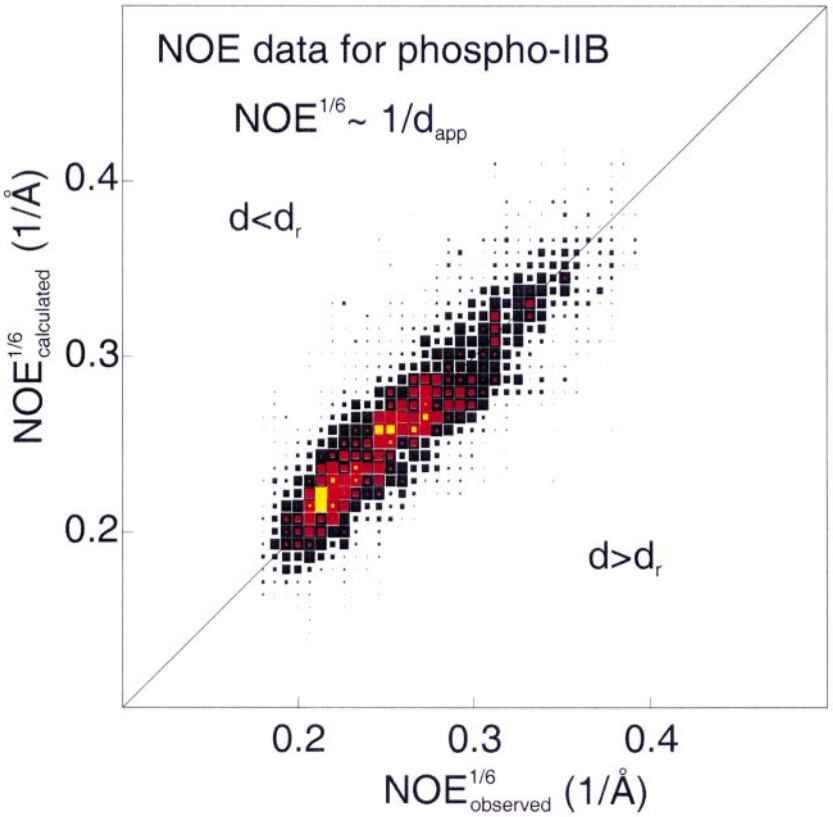

Figure 6. Comparison of measured and calculated data, for the eight selected models. Shown is the color-coded density of the sixth-root NOE intensities, equivalent to the inverse apparent distances. The area of the black, red and yellow squares is proportional to the number of restraints in a certain range. Spin diffusion corrections for the calculated intensities are calculated as described in Materials and Methods.

difference between the calculated and the observed couplings, disregarding data deviating more than 2σ. This yielded rmsd values between measured and calculated couplings of approximately 2.0-2.5 Hz. The results for the residues in the P-loop are presented in Table 3.

Discussion

From the calculated structures and the chemical shift differences we conclude that the structure of the phosphorylated form of enzyme IIB^{Chb} differs from the previously determined structures of the IIB^{Chb} wild-type[13] and the Cys10ser mutant[14] only in the active-site region. This region consists of the P-loop (residues 10-16), and the side-chains of residues Gln59 and Tyr84. Figure 7(a) shows a superposition of the backbone of the Cys10Ser mutant, of the wild-type and of the phosphorylated protein. The major secondary structure elements, a four-stranded parallel β-sheet flanked by five α-helices, show no significant changes upon phosphorylation, and can be superimposed very precisely on the X-ray structure. It was already proposed by the authors of the IIB^{Chb} X-ray structure[13] that, due to crystal contacts, the conformation of the P-loop in the X-ray structure would not be representative for the conformation in solution, but that instead it would assume a more cradle-like conformation, like in the PTPases. Indeed, as shown in Table 3, the dipolar coupling measurements show large discrepancies with the calculated values for several residues of the P-loop, when the X-ray structure was used as the model. This confirms that the structure of the P-loop of phosphorylated IIB^{Chb} in solution is significantly different from that in the X-ray structure. On the other hand, in the representative NMR structure, the individual ^{15}N-^{1}H vectors of the P-loop need a rotation of maximally 7° to bring them into agreement with the data.

The differences with the (Cys10Ser)IIB NMR structures are, in general, within the limits of accuracy of the latter, but the P-loop is much less well defined in the mutant structures. Although no relaxation or exchange data are available to give information on the mobility of the P-loop, it seems reasonable to assume that after phosphorylation the mobility of the P-loop is less than in the wild-type and Cys10Ser mutant because of the lack of suitable hydrogen bond acceptors in the latter two. Keyhani *et al*[22] found that the sedimentation coefficient increases upon phosphorylation. This indicates that upon phosphorylation, phospho-IIB^{Chb} becomes more compact or that changes in hydration occur, and agrees well with a model in which the mobile P-loop in the unphosphorylated protein becomes more defined, and/or less hydrated in the phosphorylated form.

From the structures determined thus far, there is no evidence for a conformational change (other than in the active site) that could be able to provide a conformational coupling mechanism for coupling phosphorylation and transport. Nevertheless, the chemical shift differences are, on the average, quite large compared to similar data for HPr,[23] where observable shift differences upon phosphorylation are limited to a very few residues close to the phosphoryl group. As IIB^{Chb} has a high number of aromatic amino acid residues in its core, small changes in dynamic behaviour in the vicinity of these residues might explain the relatively large chemical shift effects.

Several groups of the P-loop act in stabilizing the phosphoryl group. The backbone amide protons of Ala12, Gly13, Met14 and Ser15 are within hydrogen bonding distance of the oxygen atoms of the phosphoryl group. The amide protons of Ala12 and Ser15 probably have the strongest hydrogen bonding interactions with the phosphoryl oxygen atoms. They show a very acidic (downfield) chemical shift value in the phosphorylated form of the enzyme, which is often indicative of close interaction with a negatively charged ion. Other

Table 3. Measured and calculated dipolar couplings, *D*(Hz), for N-H bond vectors of the residues of the P-loop, for the eight selected NMR structures (WTP), the representative model (WTP-R), and the first model of the unphosphorylated wild-type X-ray structure (PDB code: 1iib)

		WTP	WTP-R		X-ray	
Residue	D_{obs}	$\langle D_{calc} \rangle$ (rms)	D_{calc}	α_{min}	D_{calc}	α_{min}
Cys10	4.2	7.6 (0.5)	7.1	5.6	5.5	2.9
Ser11	0.1	6.5 (2.5)	4.6	6.7	7.5	15.2
Ala12	−6.7	−10.6 (1.8)	−10.6	6.0	2.7	35.1
Gly13	10.6	11.6 (1.5)	12.6	6.0	−4.8	35.5
Met14	12.3	14.6 (2.0)	12.9	1.1	10.4	6.1
Ser15	-	−1.7 (4.3)	−4.9	-	3.1	-
Thr16	-	−18.3 (1.9)	−19.8	-	−3.9	-
Ser17	3.3	0.7 (1.5)	1.2	4.7	6.4	9.1
Leu18	12.9	11.1 (0.9)	10.3	5.0	11.0	5.0
Leu19	1.6	4.1 (0.9)	5.4	5.9	−4.7	10.7

The α_{min} values are the minimal angles over which the bond vectors have to be rotated to reproduce the measured dipolar couplings. For residues 15 and 16, no coupling could be observed. For residue 18, no isotropic reference value was measured.

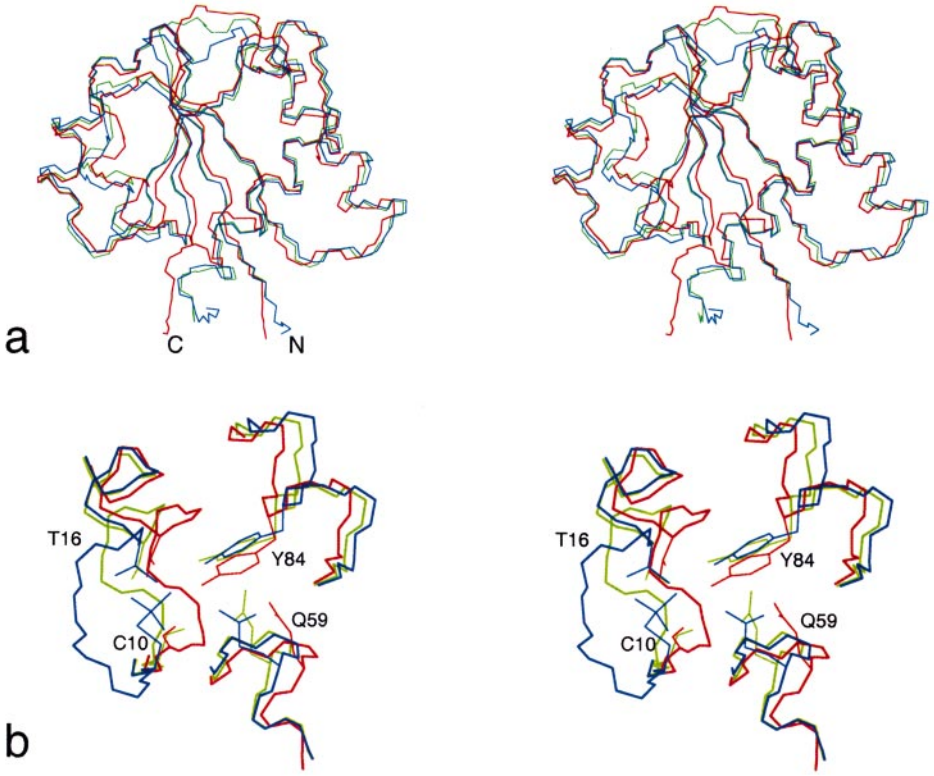

Figure 7. Comparisons of the X-ray structure of the wild-type (red), the mean NMR structure of serine mutant (green) and the representative NMR structure of the phosphorylated wild-type (blue), after superposition on the C^{α}-atoms of residue 3-10 and 17-100. (a) Backbone traces. (b) Active-site region.

hydrogen bond donors are the side-chain amide group of Gln59 and possibly the hydroxyl group of Ser15. The NH_2-group of Gln59 appears to be acting as a bridge between the phosphoryl group oxygen atoms and the hydroxyl group of Tyr84, which has moved away from the phosphoryl group when compared to the X-ray structure of the wild-type, and the NMR structure of the Cys10Ser mutant (see Figure 7(b)). The amide proton of Gly13 is probably hydrogen bonded to Cys10 S^{γ}, as indicated by the unusually short distance between these atoms. Also the γ-hydroxyl moiety of Thr16 is within hydrogen bonding distance of the Cys10 S^{γ} in a number of the structures. Table 4 summarizes the relevant distances in the eight selected structures. Figure 8 shows a drawing of the active site.

The PTPases and the dual specificity (Tyr/Thr) protein phosphatases are the only known enzymes that, like the IIB domains/protein of the PTS, use a phosphocysteine residue in the catalytic cycle. The active site of IIB^{Chb} is similar to that of the PTPases. The X-ray structure of a phosphorylated PTPase (PTP1B) has been solved, taking advantage of a Gln281Ala mutation that has a reduced dephosphorylation rate.[24] Comparison shows that the structures of the active sites of phospho-IIB^{Chb} and phospho-(Gln281Ala)PTP1B are very similar (see

Table 4. Interactions in the active site

Donor		Acceptor		Distance ⟨r⟩ (σ) (Å)	Min-Max	Δppm	Angle D-H···A
Ala12	H^N	pCys10	O^P	2.30 (0.47)	1.80-3.34	2.46	166 (7)
Gly13	H^N	pCys10	S^{γ}	2.43 (0.23)	2.03-2.84	−0.26	138 (9)
Met14	H^N	pCys10	O^P	2.24 (0.24)	1.92-2.74	−0.11	160 (8)
Ser15	H^N	pCys10	O^P	2.30 (0.17)	2.08-2.61	2.05	155 (13)
Ser15	O^{γ}	pCys10	O^P	3.22 (0.20)	2.96-3.49		
Thr16	$H^{\gamma 1}$	pCys10	S^{γ}	3.47 (0.54)	2.85-4.48		
Thr16	$O^{\gamma 1}$	pCys10	S^{γ}	3.88 (0.42)	3.40-4.78		
Gln59	$H^{\varepsilon 21}$	pCys10	O^P	2.27 (0.19)	2.06-2.61	0.39	149 (20)
Tyr84	O^{η}	pCys10	O^P	4.07 (0.52)	3.01-4.80		

Distances are measured to the closest oxygen atom of the phosphoryl group. The values in the second last column refer to the difference in ppm values of the respective amide protons between the phosphorylated enzyme and the Cys10Ser mutant.

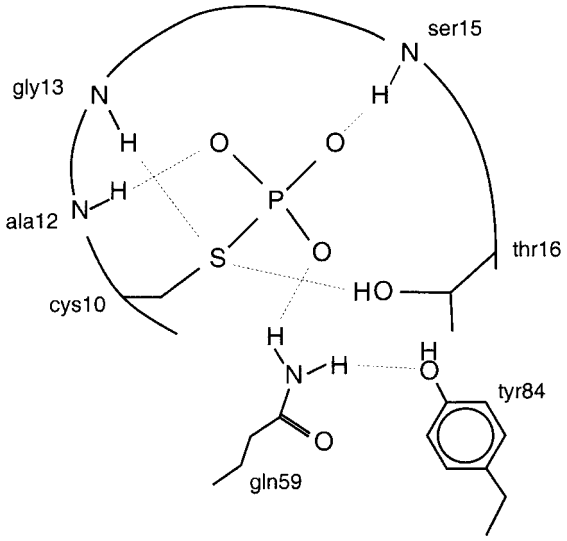

Figure 8. A drawing of the binding mode of the phosphoryl group in the active site with the hydrogen bond interactions indicated (broken lines).

Figure 9). The heavy backbone atoms, N, C^{α} and C′, of residues 10-15 and the heavy side-chain atoms of residue 10 can be superimposed with an rmsd of 0.54 Å.

In both structures, the cysteine residue is located at the end of a β-strand in a sheet-loop-helix motif. The α-helix points with its N-terminal side to the phosphoryl group, thereby stabilizing the negative charge.[25]

The γ-hydroxyl group of Ser222 (residue $i+7$ from the cysteine) of PTP1B has been proposed to polarize S^{γ} of the phosphocysteine residue in order to facilitate hydrolysis of the phosphoryl group.[24] We find that in IIB^{Chb} the hydroxyl group of the Thr at position $i+6$, is positioned in a comparable way, but with a somewhat larger distance to S^{γ}, on average. Probably the (tentative) hydrogen bond of the Cys10 side-chain S^{γ} with the Thr16 γ-hydroxyl group is weaker than the comparable hydrogen bond of Ser222 in PTPIB.

In PTP1B, the cysteine residue is assumed to be in the ionized form before phosphorylation.[24,26] There are no data available on the protonation state of the cysteine residue in IIB^{Chb}. As there are no charged groups in the neighborhood of the cysteine residue in IIB^{Chb}, unlike the PTPases and some other enzymes IIB, we assume that the cysteine residue in the free wild-type enzyme is protonated, at the pH used. In that case, the short distance between the hydroxyl group of Thr16 and S^{γ} of Cys10 (even shorter in the wild-type X-ray structure) suggests the presence of a proton relay pathway from Cys10 S^{γ} *via* Thr16 $O^{\gamma 1}$ to a solvent water molecule. Although Thr16 O^{γ} is not very solvent-exposed, it is located close to the surface and only small changes are necessary to make it accessible for a solvent molecule.

Our observation that the distance between the hydroxyl group and the cysteine S^{γ} in phospho-IIB^{Chb} is somewhat larger than in PTP1B, and the fact that Ser222 in PTP1B is not accessible to solvent, is most likely related to the hydrolysis behaviour of both enzymes. In the PTPases, the PTP1B phosphoryl group needs to be hydrolyzed after phosphorylation, while the phosphoryl group on IIB needs to be stable until it can be transferred to the carbohydrate. A larger distance in the free phospho-IIB^{Chb} could serve to increase the resistance to hydrolysis. It is feasible that the binding of IIA^{Chb} or IIC^{Chb}/carbohydrate induces a small conformational change in the active site that reduces the distance between these two groups, thereby facilitating the proton transfer or possibly, in the case that the wild-type form is ionized, stabilizing an ionized cysteine.

Modelling studies on the transition state IIA^{Chb}-P-IIB^{Chb}, indicate that the free backbone hydrogen bond acceptors of residues 11-13 in IIB^{Chb} are accessible to an arginine side-chain of IIA^{Chb} that is located close to the active histidine residue of IIA^{Chb} (R. de Jong & B.W. Dijkstra, unpublished results). This would provide a possible mechanism for IIA^{Chb} to influence the conformation of the P-loop of IIB^{Chb}.

In the PTPases, hydrolysis of the phosphoryl group starts with a solvent molecule that is activated by Asp129 (not visible in Figure 9) to make an in-line nucleophilic attack on the phosphoryl group. If we can draw a parallel between the PTPases and the enzymes IIB, we would expect the presence of a charged or polar group that is able to accept a proton from the hydroxyl group of the sugar bound by the IIB-IIC complex. As no obvious candidate is available on IIB^{Chb}, that group will probably be located in enzyme IIC.

The currently presented structure of phospho-IIB^{Chb} provides clues about the mechanism of phosphorylation and dephosphorylation of the enzyme, and can be used as a starting point for kinetic measurements and site-directed mutagenesis experiments to further unravel the mechanism.

Materials and Methods

Sample preparation

Transformation of *E. coli* strain W3110 with plasmid pJR-BLIIB, carrying the gene for IIB^{Chb} with control elements for the temperature-sensitive lambda-repressor, culturing and purification of the wild-type IIB^{Chb} was performed as described for the Cys10Ser-mutant.[27] About 20 mg of both ^{15}N and $^{13}C/^{15}N$ labeled protein was obtained.

The enzyme was phosphorylated by addition of 1 % (w/w) equivalents of enzyme I, HPr and IIA^{Chb}, in TAG-buffer (10 mM Tris-acetate (pH 7.0), 5 % (w/v) glycerol, 100 mM NaCl, 1 mM NaN_3) with 5 mM DTT, 5 mM $MgCl_2$ and 13.3 mM mono(cyclohexyl-ammonium)-PEP. Phosphorylation of IIB^{Chb} was confirmed by the observation of a shift in the major protein band using non-denaturing PAGE.

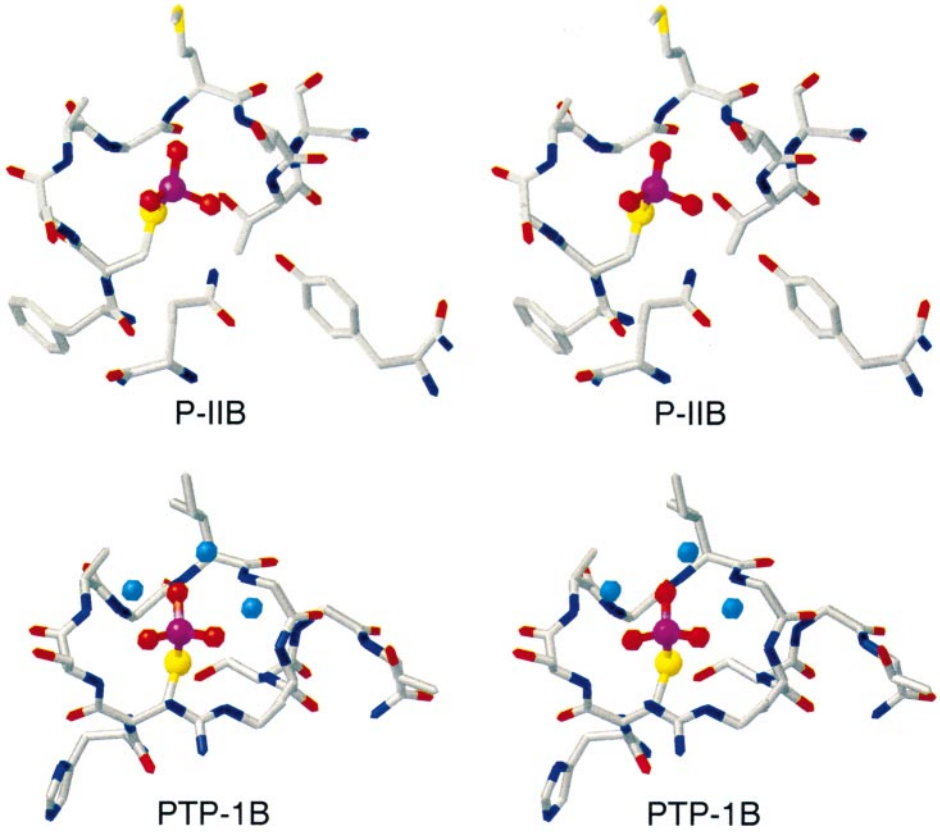

Figure 9. Stereoviews of the active sites of phospho-IIB^{Chb} (P-IIB) and PTP-1B Q281A (PTP1B, PDB code: A5Y.PDB). The Figure was made using the program MOLMOL.[63] Three water molecules in the PTP1B active site are colored light blue.

The phosphoryl group of P-IIB^{Chb} is sensitive to hydrolysis under the experimental conditions used and, as a consequence, during the entire span of the experiment pyruvate is being produced, whilst PEP is being consumed. The release of pyruvate lowers the pH of the sample, causing shifts in resonances, which in turn can result in severe artifacts in the spectra. We chose a relatively high buffer concentration compared to the NMR work on the Cys10Ser mutant to reduce these effects as much as possible.

The phosphorylation reactions were started by the addition of a PEP solution that was brought to pH 7.0 by the addition of $NaHCO_3$. Complete phosphorylation of IIB^{Chb}, as judged from the spectra, was maintained during the recording of the NMR spectra for approximately one week at 295 K under the conditions listed below. In some cases the samples were regenerated by dialysing against fresh buffer without enzymes and PEP.

^{15}N and ^{13}C-^{15}N-labeled NMR samples

The composition of the samples was: (1) ^{15}N-labeled sample, 23 μM enzyme I; 34 μM HPr; 20 μM IIA^{Chb}; 4 mM IIB^{Chb}; 100 mM PEP; 100 mM TA (pH 7.0); 100 mM NaCl; 5 mM DTT; 4 mM $MgCl_2$; 1.5% (w/v) Glycerol-D5. (2) The ^{13}C/^{15}N-labeled sample, as (1) except 80 mM NaCl. (3) ^{13}C/^{15}N-labeled sample, as (2) except 100 mM PEP (4) as (3) except: 2 mM IIB^{Chb}; 30 mM PEP; molar ratio C835 5% (w/v); octanol 1.6% (w/v).

NMR spectroscopy

All measurements were carried out on a Varian Inova 600 MHz spectrometer equipped with a pulsed-field gradient probe. The spectra were recorded at 295 K.

An unlabeled sample was used to acquire a 2D-NOESY spectrum. The ^{15}N-NOESY-HSQC and ^{15}N-TOCSY-HSQC[28–30] were performed as described[27] on ^{15}N-labeled protein. Pulsed-field gradients were added to reduce phase-cycling. A ^{13}C,^{15}N-labeled sample was used to acquire the following 3D triple resonance experiments: CBCA(CO)NH, HBHA(CBCA)(CO)NH, CBCANH, CANH, HNCO, HN(CA)CO, COCAH, (CO)N(CO)CAH, HCACBCO, $C^{\beta}H^{\varepsilon}$ and $C^{\beta}H^{\delta}$ (for aromatic ring protons), HCCH-TOCSY, ^{13}C-HSQC-NOESY, high-resolution constant-time ^{13}C-HSQC. Additionally. $^{1}J_{NH}$-modulated HSQC spectra at 288 K and at 298 K were recorded to determine residual dipolar couplings in oriented C8E5/octanol medium[31] The relevant parameters and references for the acquired spectra are listed in Table 5.

Resonance assignments

We performed a considerable number of the side-chain assignments during the structure calculations, on the basis of the NOE spectra, because the other spectra were insufficient to obtain the complete assignments of side-chain resonances. Numerous cycles of structure calculations, resonance assignments and re-evaluation of the assignments were performed. Details on various

Table 5. NMR-experiments used for the resonance assignments and structure determination of phospho-IIB^{Chb}

Spectrum	Nucl.[a]	Spec. width[b] (kHz)	Carrier (ppm)	Max. evol. time (ms)	Method	[c]	References/remarks
^{15}N-HSQC	^{15}N	2.5	117.082	204.7	TPPI	1	Bodenhausen & Ruben[43]
	^{1}H	8.0	4.830	128.0			
^{15}N-NOESY-HSQC	^{1}H	8.0	4.829	12.0	TPPI	1	Marion *et al.*[28]
	^{15}N	2.5	115.442	51.2	States		Zuiderweg & Fesik[29]
	^{1}H	8.0	4.829	128.0			Fesik & Zuiderweg[30]
^{15}N-TOCSY-HSQC	^{1}H	8.0	4.829	12.0	TPPI	1	Bax & Davis[44]
	^{15}N	2.5	117.087	48.8	States		Marion *et al.*[45]
	^{1}H	8.0	4.829	128.0			
CBCA(CO)NH	$^{13}C^{\alpha\beta}$	12.0	45,680	5.33	TPPI	2	Grzesiek & Bax[46]
	^{15}N	3.0	117.082	22.33	TPPI		
	^{1}H	8.0	4.830	64.0			
CBCANH	$^{13}C^{\alpha\beta}$	12.0	55.629	5.33	TPPI	2	Grzesiek & Bax[47]
	^{15}N	3.0	117.085	22.33	TPPI		
	^{1}H	8.2	4.838	64.0			
CANH	$^{13}C^{\alpha}$	5.0	56.292	12.8	TPPI	2	Kay *et al.*[48]
	^{15}N	2.5	117.086	25.2	TPPI		
	^{1}H	8.0	4.838	63.9			
HNCO	$^{13}C'$	3.0	176.471	21.3	TPPI	2	Grzesiek & Bax[47]
	^{15}N	2.5	117.086	24.8	TPPI		Kay *et al.*[50]
	^{1}H	8.0	4.833	64.0			
HN(CA)CO	$^{13}C'$	3.0	58.030	21.3	TPPI	2	Clubb *et al.*[51]
	^{15}N	2.5	117.086	24.8	TPPI		
	^{1}H	8.0	4.833	64.0			
HBHA(CBCA)(CO)NH	$^{1}H^{\alpha\beta}$	4.0	3.166	16.0	TPPI	2	Grzesiek & Bax[52]
	^{15}N	3.0	117.085	42.7	TPPI		
	^{1}H	8.0	64.0	42.7			
COCAH	$^{13}C'$	3.0	177.351	12.5	TPPI	2	Dijkstra *et al.*[53]
	$^{13}C^{\alpha}$	6.0	58.002	5.42	TPPI		
	^{1}H	8.0	4.833	64.0			
(CO)N(CO)CAH	^{15}N	2.4	117.100	18.96	STPPI	2	Dijkstra *et al.*[54]
	$^{13}C^{\alpha}$	6.0	58.002	5.25	TPPI		
	^{1}H	8.0	4.833	64.0			
HCACBCO	^{13}C	12.0	42.366	5.33	TPPI	2	Kay[55]
	$^{13}C'$	3.0	178.216	42.67	TPPI		
	^{1}H	8.0	4.833	63.93			
HCCH-TOCSY	^{1}H	5.0	2.545	19.2	TPPI	2	Olejniczak *et al.*[56]
	^{13}C	12.0	45.683	10.6	TPPI		PFG-enhanced
	^{1}H	5.0	2.545	102.5			Kay *et al.*[52]
2D-NOESY	^{1}H	8.0	4.813	64.0	States	1	τ_{mix}=100 ms
	^{1}H	8.0	4.813	127.9			
^{13}C-HSQC-NOESY	^{1}H	8.0	4.833	12.0	STPPI	3	Majumdar & Zuiderweg[58]
	^{13}C	13.0	40.307	7.346	TPPI		τ_{mix}=100 ms
	^{1}H	8.0	4.833	64.0			
$C^{\beta}G^{\delta}/C^{\beta}H^{\varepsilon}$	^{13}C	6.0/5.0	35.733	8.0	States	3	Yamazaki *et al.*[59]
	^{1}H	8.0	4.830	63.9			
$^{1}J_{NH}$-modulated HSQC	^{15}N	2.5	117.163	102.4	STPPI	4	Tjandra *et al.*[60]
	^{1}H	8.0	5.083	127.9			

[a] Spectral domains are in order ω1, ω2, (ω3).
[b] The carrier position is given in ppm relative to DSS for ^{1}H and ^{13}C, and relative to liquid NH_3 for ^{15}N.
[c] Sample number, see Materials and Methods.

aspects of the generation of restraints and structure calculations are outlined in the following paragraphs.

NOE assignments

The cross-peaks for each NOE spectrum were divided into two lists according to whether the peaks were assigned to a specific resonance (list a) or not (list b). For list b, all assignments that were possible on the basis of the assigned chemical shift lists were generated. Acceptance margins for the NOE assignments were based on the spread in the assigned shifts in the spectrum or spectra, and on the digital resolution of the spectral data. Subsequently the list of possible assignments was cleaned up by removing all the pairs with distances larger than 8.0 Å in a reference set of structures. Assignments that were already present in list a were removed from list b. In the first stages of calculations, the NMR structures of the Cys10Ser mutant and the X-ray structures of the wild-type were used as reference set, in later stages the structures obtained from previous rounds of structure calculations were used.

Generation of distance restraints

For the generation of distance restraints, we used various protocols depending on the stage of refinement and possibilities of the software used.

For conversion of NOE-intensities$I$ (peak maxima) to restraint distances R we used the relation $R \propto I^{-1/6}$, where the scaling factor is chosen such that the average ratio between apparent model and observed distances

equals one: $\langle R^{app}_{calc}/R^{app}_{obs}\rangle = 1$. The calibration was done with the unambiguously assigned NOEs. From the scaled distances, upper (and lower) bounds were generated, in a manner depending on the structure calculation or refinement program used.

Spin diffusion corrections

Restraint distances were corrected for multi-spin effects, ''spin diffusion'',[32] by comparing the apparent distances in a reference model for a given restraint with and without spin-diffusion correction. Apparent distances, taking into account spin diffusion, were calculated as described by Kemmink & Scheek[33]

The rotational correlation time was estimated from plots of the observed data against the data calculated using different values of the correlation time. The standard deviation of the ratio NOE_{obs}/NOE_{calc} for all NOEs was smallest at a value of 11 ns. This value of the correlation time agrees well with the rule-of-thumb that says that $\tau \approx M_r$ (ns/kDa).

The spin diffusion correction (SDC) that was applied equals the difference between the two apparent distances: $SDC = R^{spindif}_{calc} - R_{calc}$, where the R values are the inverse sixth-power summed distances for a set of spin pairs.

If a cluster of structures was used to calculate the corrections (for the embedding and DDD stages), the highest correction distance found in the structures (SDC_{max}) was added to the upper bound and the lowest correction distance (SDC_{min}) to the lower bound, in order to reflect the widest possible bounds compatible with the reference models. If one structure was used to calculate the corrections (for the refinement using GROMACS), the correction was straightforwardly added to the restraint distance (upper and/or lower bound).

Use of ambiguous distance restraints

For the use of ambiguous restraints,[34] including the treatment of prochiral ambiguities, we used the approach as explained by AB *et al.*,[14] with some modifications. Upper and lower bounds were generated for all unambiguously assigned NOE peaks. From the NOE intensities that were not yet assigned unambiguously, ambiguous restraints with only upper bounds were generated. Aromatic ring resonances from the ^{13}C-HSQC-NOESY were excluded from this procedure, as they showed complicated multiplet-like signals because of incomplete decoupling of the ^{13}C nuclei during acquisition. This resulted in cases where several peaks in these multiplets were assigned to the same resonance pair. In these cases the added intensities have to be considered as lower estimates for the total intensity. Consequently, from these intensities only upper bounds were generated.

Due to crowding of resonances in the spectra, or uncertainties in the frequency assignments, it occurred that some spin pairs appeared as possible assignments for more than one NOE peak. In this situation we speak of ''assignment overlap'' and the assignments we call ''shared assignments''. By using the standard approach, in which the possible assignments for each peak are simply used as pairs in an ambiguous restraint, we lose information, because only one close pair, namely that of the shared assignment, is needed to satisfy all the restraints that share that pair. To prevent information loss in these cases we applied the following procedure. (1) We added an ambiguous restraint corresponding to the sum of the intensities to the list, containing as pairs all the possible assignments of the original peaks. (2) Furthermore, we added a restraint corresponding to the lowest intensity, with all possible assignments of the original peak except the shared assignment. These extra restraints ensure that (1) the sum of intensities of two NOEs with shared assignments is also restrained, instead of just the highest intensity, and (2) at least two different pairs are necessary to satisfy the two NOE intensities. Further complications due to assignments shared by more than two peaks have also been treated. Further details will be provided elsewhere (E.A. & R.M.S., unpublished results).

After generating additional combined restraints to deal with shared assignments for each NOE spectrum separately and adding the resulting restraint lists, redundant restraints were removed from the list.

Treatment of prochiral ambiguities

In the DDD stage with 3D coordinates, prochiralities were handled by using the flipping chiralities approach as described by van Nuland *et al.*[35] For the refinement of structures in explicit solvent using GROMACS, prochiralities were handled by the construction of appropriate combinations of ambiguous distance restraints, as explained by AB *et al.*[14] (and E.A. & R.M.S., unpublished results) in order to make the restraints insensitive to the interchange of prochiral groups.

Stereospecific assignments of prochiral pairs were done if they were consistent for all calculated structures after the stages of embedding and DDD in 4D and in 3D. For use in GROMACS, non-assigned prochiralities were expanded into ambiguous restraints, and subsequently, together with the other restraints, expanded into appropriate sums and subsets. After distance restrained simulated annealing in GROMACS, the different possible assignments for each prochiral center were compared. Per-residue violations and R-factors were calculated for each possible prochiral assignment. If one assignment consistently showed lower violations or R-factors for all structures, the assignment was fixed in the next round of calculations. The following R-factor was used:

$$R_2^x = \frac{\Sigma(D_{obs}^{-1} - D_{calc}^{-1})^2}{\Sigma(D_{obs}^{-1})^2}$$

where D_{obs} and D_{calc} are the apparent model and experimental distances and the data was scaled such that $\langle D_{calc}/D_{obs}\rangle = 1$.

Structure calculations

Structure calculations were done in three stages. (1) All unambiguous restraints were used in a distance geometry (DG) procedure to create 4D coordinates. (2) In order to take advantage of all available unambiguous and ambiguous NOE data to resolve prochiral ambiguities, we performed distance driven dynamics (DDD),[14,36–38] with a high-temperature search phase in 4D, where the energy required to invert chiralities is strongly reduced compared to 3D. Subsequent DDD was done to anneal the structures to 3D coordinates. (3) In order to improve the stereochemical quality, we subjected the structures to further refinement in a physically more realistic force field in explicit solvent, using the

molecular dynamics software GROMACS.[39] Three separate sets of restraints suitable for these structure calculation and refinement protocols were prepared. Distance restraints for 54 hydrogen bonds were defined.

Distance geometry and distance driven dynamics

The unambiguous upper and lower bounds, corrected for spin diffusion, were collected in a matrix. Distances of $0.05D + 0.05D^2/\text{Å}$ and $0.05D - 0.05D^2/\text{Å}$ were added to and subtracted from the upper and lower bounds, respectively. The lower bounds were adjusted to make sure that the difference between the upper and lower bound was at least 1.25 Å. Additional bounds were calculated using triangule inequalities. A submatrix of this smoothed matrix, consisting of all the non-hydrogen atoms and the amide and methine hydrogen atoms, was used to calculate 4D coordinates using the EMBED algorithm.[38,40] The distance driven dynamics approach was used to refine the structures further. The error function consisted of all ambiguous and unambiguous restraints and hydrogen bonds, as well chiral restraints that impose the correct handedness for asymmetric carbon atoms and α-helices. The lower bounds were readjusted to have at least 0.5 Å difference with the upper bounds. One thousand steps of dynamics in 4D were performed at 1000 K. During the last 500 steps a force was applied to minimize the fourth coordinate. Subsequently, simulated annealing was done by temperature coupling to an external thermal bath of 500, 300, 200 and 1 K, for 100 steps each.[41]

After projection of the 4D coordinates into 3D space,[14] the 3D coordinates were subjected to further conjugate-gradient optimization and DDD calculations, using the same error function in a protocol similar to that described above.

Robust DDD

In the earlier stages of the refinement, there was still a considerable number of errors in the spectral assignments and consequently, in the restraints. We modified the DDD algorithm in order to make it less sensitive to large errors. During each time-step, the n largest violations were determined, where n is typically 10, and the force resulting from these violations was either discarded, or scaled down typically five to tenfold. In this way we allowed for the restraint list to contain a number of errors, without the structures being severely distorted by them. Large violations in the resulting structures were subsequently used as guidance for manual re-evaluation of the assignments.

GROMACS

In order to improve the quality of the structures we used GROMACS to perform simulated annealing in a physically more realistic force-field with explicit solvent. For these runs, we used calibrated and spin-diffusion corrected tight bounds, $D_{up} = D_{lo} = D_r$, with a relatively low force constant for restraining. The restraints follow the following potential function: $E = (D_o/D_r) \times 1/2 \times k_{dr} (D - D_r)^2$, where $D_o = 0.3$ nm and $k_{dr} = 2000$ kJ mol nm^2. For values of D larger than $D'_r = D_r + q/D_r$ ($q = 0.04$ nm^2) a linear potential energy function was used. For restraints with an upperbound of 3.0 and 5.0 Å a violation of 1.0 Å corresponds to an energy of 10 and 3.84 kJ mol^{-1}, respectively, while the maximum forces in these cases are reached at energies of 17.78 and 3.84 kJ mol^{-1}, respectively. For both ambiguous and unambiguous restraints, forces were applied proportional to the derivative of the potential function. In order to increase the possible time-step, we used a united atom force-field that was modified to include dummy hydrogen atoms relative to which the restraint forces were defined.†[42]

The refinement runs consisted of a high-temperature search phase starting at 500 K and a cooling phase, linearly to 300 K in 7.2 ps with time-steps of 4 fs, followed by a cooling stage from 300 K to 0 K in 3.2 ps with time-steps of 6 fs. The runs were repeated several times and the spin diffusion corrections were recalculated before each run from the results of the previous run. Usually sixth-root R-factors[19] had converged to stable values after three rounds.

Selection of structures

Because the conformations of the active-site residues was most relevant to this study, we selected the structures that showed the smallest per residue R_2^x value for the active-site residues 10-16, 59, 84.

To obtain a high-quality representative model, we applied the following procedure. For each residue, the coordinates of the residue, with the lowest per-residue R_2^x value in the superimposed 32 structure bundle was taken. The resulting (WTP-R) structure was energy minimized in the GROMACS force-field, with NOE restraints. To judge whether the structure converged to low energies, we subsequently subjected it to 55 steps of steepest descent minimization (WTP-F) without restraints.

Data Bank accession numbers

The chemical shifts are deposited at the BioMag ResBank‡, BMRB Accession number 4955.

The protein structures have been deposited at the Protein Data Bank, PDB code 1h9c.

Wild-type and Cys10Ser mutant structures can be found at PDB with codes 1e2b and 1iib, respectively.

Acknowledgements

We thank Berk Hess and Anton Feenstra for implementing the distance restraining and dummy atom options in GROMACS, and Marc Lensink for providing the phosphocysteine bond lengths and angles. Frans van Hoesel is gratefully acknowledged for writing SNARF. We thank Franciska van Lune for assistance with the dipolar coupling measurements. Unfortunately one of the prospected co-authors, Jonathan Reizer, who created the overproduction vectors with which this work was carried out, passed away before this manuscript was finished. We are indebted to him for his contributions to this work. Jonathan was an enthusiastic, creative scientist who seemed to live solely for science. His death is a great loss to those who have known him personally or followed his scientific endeavors.

† GROMACS User Manual version 2.0 can be found at http:md.chem.rug.nl/~gnx

‡ http://www.bmrb.wisc.edu

References

1. Parker, L. L. & Hall, B. G. (1990). Characterization and nucleotide sequence of the cryptic cel operon of *Escherichia coli* K12. *Genetics,* **124**, 455-471.
2. Reizer, J., Reizer, A. & Saier, M. H., Jr (1990). The cellobiose permease of *Escherichia coli* consists of three proteins and is homologous to the lactose permease of *Staphylococcus aureus*. *Res. Microbiol.* **141**, 1061-1067.
3. Keyhani, N. O. & Roseman, S. (1997). Wild-type *Escherichia coli* grows on the chitin disaccharide, *N*,*N*′-diacetylchitobiose, by expressing the cel operon. *Proc. Natl Acad. Sci. USA,* **94**, 14367-14371.
4. Postma, P. W., Lengeler, J. W. & Jacobson, G. R. (1993). Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* **57**, 543-594.
5. Meadow, N. D., Fox, D. K. & Roseman, S. (1990). The bacterial phosphoenolpyruvate:glycose phosphotransferase system. *Annu. Rev. Biochem.* **59**, 497-542.
6. Lolkema, J. S., Swaving-Dijkstra, D. & Robillard, G. T. (1992). Mechanics of solute translocation catalyzed by enzyme IImtl of the phosphoenolpyruvate-dependent phosphotransferase system of *Escherichia coli*. *Biochemistry,* **31**, 5514-5521.
7. Saier, M. H., Jr & Reizer, J. (1992). Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *J. Bacteriol.* **174**, 1433-1438.
8. Robillard, G. T. & Broos, J. (1999). Structure/function studies on the bacterial carbohydrate transporters, enzymes II, of the phosphoenolpyruvate-dependent phosphotransferase system. *Biochim. Biophys. Acta,* **1422**, 73-104.
9. Lolkema, J. S., Swaving-Dijkstra, D., ten Hoeve-Duurkens, R. H. & Robillard, G. T. (1991). Interaction between the cytoplasmic and membrane-bound domains of enzyme IImtl of the *Escherichia coli* phosphoenolpyruvate-dependent phosphotransferase system. *Biochemistry,* **30**, 6721-6726.
10. Lolkema, J. S., ten Hoeve-Duurkens, R. H., Swaving-Dijkstra, D. & Robillard, G. T. (1991). Mechanistic coupling of transport and phosphorylation activity by enzyme IImtl of the *Escherichia coli* phosphoenolpyruvate-dependent phosphotransferase system. *Biochemistry,* **30**, 6716-6721.
11. Meijberg, W., Schuurman-Wolters, G. K. & Robillard, G. T. (1998). Thermodynamic evidence for conformational coupling between the B and C domains of the mannitol transporter of *Escherichia coli*, enzyme IImtl. *J. Biol. Chem.* **273**, 7949-7956.
12. Keyhani, N. O., Bacia, K. & Roseman, S. (2000). The transport/phosphorylation of *N*,*N*′-diacetylchitobiose in *Escherichia coli*. Characterization of phospho-IIBChb and of a potential transition state analogue in the phosphotransfer reaction between the proteins IIAChb and IIBChb. *J. Biol. Chem.* **275**, 33102-33109.
13. van Montfort, R. L. M., Pijning, T., Kalk, K. H., Reizer, J., Saier, M. H., Jr, Thunnissen, M. M. G. M., Robillard, G. T. & Dijkstra, B. W. (1997). The structure of an energy-coupling protein from bacteria, IIBcellobiose, reveals similarity to eukaryotic protein tyrosine phosphatases. *Structure (London),* **5**, 217-225.
14. AB, E., Schuurman-Wolters, G. K., Reizer, J., Saier, M. H., Jr, Dijkstra, K., Scheek, R. M. & Robillard, G. T. (1997). The NMR side-chain assignments and solution structure of enzyme IIBcellobiose of the phosphoenolpyruvate-dependent phosphotransferase system of *Escherichia coli*. *Protein Sci.* **6**, 304-314.
15. Guan, K. & Dixon, J. E. (1991). Evidence for protein-tyrosine-phosphatase catalysis proceeding *via* a cysteine-phosphate intermediate. *J. Biol. Chem.* **266**, 17026-17030.
16. Cho, H., Krishnaraj, R., Kitas, E., Bannwarth, W., Walsh, C. T. & Anderson, K. S. (1992). Isolation and structural elucidation of a novel phosphocysteine intermediate in the LAR protein tyrosine phosphatase enzymatic pathway. *J. Am. Chem. Soc.* **114**, 7296-7298.
17. Ishibashi, T., Bottaro, D. P., Chan, A., Miki, T. & Aaronson, S. A. (1992). Expression cloning of a human dual-specificity phosphatase. *Proc. Natl Acad. Sci. USA,* **89**, 12170-12174.
18. Zhou, G., Denu, J. M., Wu, L. & Dixon, J. E. (1994). The catalytic role of Cys124 in the dual specificity phosphatase VHR. *J. Biol. Chem.* **269**, 28084-28090.
19. Thomas, P. D., Basus, V. J. & James, T. L. (1991). Protein solution structure determination using distances from two-dimensional nuclear Overhauser effect experiments: effect of approximations on the accuracy of derived structures. *Proc. Natl Acad. Sci. USA,* **88**, 1237-1241.
20. Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check to stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
21. Tjandra, N. & Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science,* **278**, 1111-1114.
22. Keyhani, N., Rodgers, M. E., Demeler, B., Hansen, J. C. & Roseman, S. (2000). Analytical sedimentation of the IIAChb and IIBChb proteins of the *Escherichia coli N*,*N*′-diacetylchitobiose phosphotransferase system. Demonstration of a model phosphotransfer transition state complex. *J. Biol. Chem.* **275**, 33110-33115.
23. van Nuland, N. A. J., Boelens, R., Scheek, R. M. & Robillard, G. T. (1995). High-resolution structure of the phosphorylated form of the histidine- containing phosphocarrier protein HPr from *Escherichia coli* determined by restrained molecular dynamics simulations. *J. Mol. Biol.* **246**, 180-193.
24. Pannifer, A. D., Flint, A. J., Tonks, N. K. & Barford, D. (1998). Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by X-ray crystallography. *J. Biol. Chem.* **273**, 10454-10462.
25. Hol, W. G. J. (1985). The role of the α-helix dipole in protein function and structure. *Prog. Biophys. Mol. Biol.* **45**, 149-195.
26. Zhang, Z. Y. & Dixon, J. E. (1993). Active site labeling of the Yersinia protein tyrosine phosphatase: the determination of the pKa of the active site cysteine and the function of the conserved histidine 402. *Biochemistry,* **32**, 9340-9345.
27. AB, E., Schuurman-Wolters, G. K., Saier, M. H., Jr, Reizer, J., Jacuinod, M. & Roepstorff, P. (1994). Enzyme IIBcellobiose of the phosphoenolpyruvate-dependent phosphotransferase system of *Escherichia coli*: backbone assignment and secondary structure determined by three-dimensional NMR spectroscopy. *Protein Sci.* **3**, 282-290.
28. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989). Overcoming the overlap problem in the

assignment of ^{1}H-NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn multiple quantum coherence and nuclear-Overhauser multiple quantum coherence spectroscopy: application to interleukin 1β. *Biochemistry,* **28**, 6150-6156.
29. Zuiderweg, E. R. P. & Fesik, S. W. (1989). Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry,* **28**, 2387-2391.
30. Fesik, S. W. & Zuiderweg, E. R. P. (1990). Heteronuclear three-dimensional NMR spectroscopy of isotopically labeled macromolecules. *Quart. Rev. Biophys.* **23**, 97-131.
31. Rückert, M. & Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793-7797.
32. Kalk, A. & Berendsen, H. J. C. (1976). Proton magnetic relaxation and spin diffusion in proteins. *J. Magn. Reson.* **24**, 343-366.
33. Kemmink, J. & Scheek, R. M. (1995). Dynamic modelling of a helical peptide in solution using NMR data: multiple conformations and multi-spin effects. *J. Biomol. NMR,* **6**, 33-40.
34. Nilges, M. (1993). A calculation strategy for the structure determination of symmetric dimers by 1H-NMR. *Proteins: Struct. Funct. Genet.* **17**, 297-309.
35. van Nuland, N. A. J., van Dijk, A. A., Dijkstra, K., van Hoesel, F. H. J., Scheek, R. M. & Robillard, G. T. (1992). Three-dimensional ^{15}N-H-H and ^{15}N-^{13}C-H nuclear-magnetic resonance studies of HPr a central component of the phosphoenolpyruvate-dependent phosphotransferase system from *Escherichia coli*. Assignment of backbone resonances. *Eur. J. Biochem.* **203**, 483-491.
36. Kaptein, R., Boelens, R., Scheek, R. M. & van Gunsteren, W. F. (1988). Protein structures from NMR. *Biochemistry,* **27**, 5389-5395.
37. Scheek, R. M., van Gunsteren, W. F. & Kaptein, R. (1989). Molecular dynamics simulation techniques for determination of molecular structures from nuclear magnetic resonance data. *Methods Enzymol.* **177**, 204-218.
38. van Nuland, N. A. J., Grötzinger, J., Dijkstra, K., Scheek, R. M. & Robillard, G. T. (1992). Determination of the three dimensional solution structure of the histidine-containing phosphocarrier protein HPr from *Escherichia coli* using multidimensional NMR spectroscopy. *Eur. J. Biochem.* **210**, 881-891.
39. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. (1995). GROMACS: a message-passing parallel molecular dynamics implementation. *Comp. Phys. Commun.* **91**, 43-56.
40. Crippen, G. M. & Havel, T. F. (1978). Stable calculation of coordinates from distance information. *Acta Crystallog.* **34**, 282-284.
41. Berendsen, H. J. C., Postma, J. P. M., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **23**, 3684-3690.
42. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. (1999). Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comp. Chem.* **20**, 786-798.
43. Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance nitrogen 15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters,* **69**, 185-189.
44. Bax, A. & Davis, D. G. (1985). MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355-360.
45. Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989). Rapid recording of 2D spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393-399.
46. Grzesiek, S. & Bax, A. (1992). Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291-6293.
47. Grzesiek, S. & Bax, A. (1992). An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99**, 201-207.
48. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **89**, 469-514.
49. Grzesiek, S. & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**, 432-440.
50. Kay, L. E., Xu, G. Y. & Yamazaki, T. (1994). Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. *J. Magn. Reson. ser. A,* **109**, 129-133.
51. Clubb, R. T., Thanabal, V. & Wagner, G. (1992). A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue proton $^{1}H^{N}$ ^{15}N and ^{13}C′ chemical shifts in ^{15}N-^{13}C′ labeled protons. *J. Magn. Reson.* **97**, 213-217.
52. Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly ^{13}C/^{15}N-enriched proteins. *J. Biomol. NMR,* **3**, 185-204.
53. Dijkstra, K., Kroon, G. J. A., van Nuland, N. A. J. & Scheek, R. M. (1994). The COCAH experiment to correlate intraresidue carbonyl, C^{α}, and H^{α} resonances in proteins. *J. Magn. Reson. ser. A,* **107**, 102-105.
54. Dijkstra, K., Kroon, G. J. A., AB, E., Scheek, R. M. & Kemmink, J. (1997). Correlation of the $^{15}N(i+1)$, $^{13}C^{\alpha}(i)$, and $^{1}H^{\alpha}(i)$ backbone resonances in ^{13}C/^{15}N-labeled proteins by the (CO)N(CO)CAH experiment. *J. Magn. Res.* **125**, 149-152.
55. Kay, L. E. (1993). Pulsed-field gradient-enhanced three-dimensional NMR experiment for correlating ^{13}C/, ^{13}C and ^{1}H chemical shifts in uniformly ^{13}C-labeled proteins. *J. Am. Chem. Soc.* **115**, 2055-2057.
56. Olejniczak, E., Xu, R. X. & Fesik, S. W. (1992). A 4D HCCH-TOCSY experiment for assigning the side-chain proton and carbon-13 resonances of proteins. *J. Biomol. NMR,* **2**, 655-659.
57. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R. & Forman-Kay, J. (1993). A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlations in H_2O samples of proteins. *J. Magn. Reson. ser. B,* **101**, 333-337.
58. Majumdar, A. & Zuiderweg, E. R. P. (1993). Improved carbon-13-resolved HSQC-NOESY spectra in water using pulsed field gradients. *J. Magn. Res. ser. B,* **102**, 242-244.
59. Yamazaki, T., Forman-Kay, J. D. & Kay, L. E. (1993). Two-dimensional NMR experiments for correlating $^{13}C^{\beta}$ and $^{1}H^{\delta/\varepsilon}$ chemical shifts of aromatic residues in ^{13}C-labeled proteins *via* scalar couplings. *J. Am. Chem. Soc.* **115**, 11054-11055.
60. Tjandra, N., Grzesiek, S. & Bax, A. (1996). Magnetic field dependence of nitrogen-proton J splittings in

^{15}N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J. Am. Chem. Soc.* **118**, 6264-6272.
61. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR,* **8**, 477-86.
62. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51-55.

Edited by P. E. Wright

(Received 8 December 2000; received in revised form 13 March 2001; accepted 14 March 2001)